USING VBAP-DERIVED PANNING FUNCTIONS FOR 3D AMBISONICS DECODING

Johann-Markus Batke and Florian Keiler

Technicolor, Research and Innovation
Karl-Wiechert-Allee 74, 30625 Hannover, Germany
{jan-mark.batke|florian.keiler}@technicolor.com

ABSTRACT

In this contribution we will explain the derivation of panning functions from the vector base amplitude panning (VBAP) technology for 3-dimensional Ambisonics decoding. VBAP turned out to be a robust method for presenting virtual acoustic sources. Each virtual source to be played back using VBAP has to be described by a monophonic signal and a 3-dimensional position as parameter. In opposite to this, Ambisonics describes an entire sound field where no explicit sources are given. This contribution shows how to apply the VBAP methodology to Ambisonics signals. In particular, we present the mathematical derivation to calculate a decoding matrix from given panning functions. Informal listening tests show that the proposed decoding approach shows much better localisation than Ambisonics decoding using mode matching for nonuniform loudspeaker setups.

1. INTRODUCTION

An accurate localisation is a key goal for any spatial audio reproduction system. Such reproduction systems are highly applicable for conference systems, games, or other virtual environments that benefit from 3D sound. Sound scenes in 3D can be synthesised or captured as a natural sound field. To synthesise audio scenes, panning functions that refer to the spatial loudspeaker arrangement, are required to obtain a spatial localisation of the given sound source. If a natural sound field should be recorded, microphone arrays are required to capture the spatial information. The playback of such captured content is a challenging task, and the Ambisonics approach is a very suitable tool to accomplish it. Ambisonics signals carry a representation of the desired sound field, and a decoding process is required to obtain the individual loudspeaker signals. Since also in this case panning functions can be derived from the decoding functions, the panning functions are the key issue to describe the task of spatial localisation.

The spatial arrangement of loudspeakers is referred to as *loudspeaker setup* in this paper. Commonly used loudspeaker setups are the stereo setup employing two loudspeakers, the standard surround setup using five loudspeakers and extensions of the surround setup using more than five loudspeakers. These setups are well known. However, they are restricted to two dimensions, e.g. no height information is reproduced. In opposite, loudspeaker setups for 3D playback are seldom described. Examples include the proposal for the NHK ultra high definition TV with the 22.2 format, the 2+2+2 arrangement of Dabringhaus and the 10.2 setup of Holman [1–3]. Only a few authors refer to spatial playback and panning strategies. One of the best known systems is the vector base amplitude panning (VBAP) approach by Pulkki [4].

The 3D loudspeaker setup example considered here has 16 loudspeakers as shown in Figure 1. The positioning was chosen due to practical considerations, having four columns with three loudspeakers each and additional loudspeakers between these columns. In more detail, eight of the loudspeakers are equally distributed on a circle around the listener's head, enclosing angles of 45 degrees. Additional four speakers are located at the top and the bottom, enclosing azimuth angles of 90 degrees. With regard to Ambisonics this setup is irregular and leads to problems in decoder design [5].

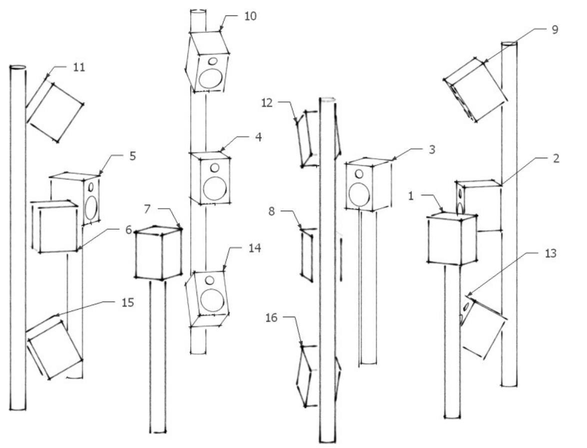


Figure 1: The example 3D setup with 16 loudspeakers as used in this paper.

VBAP is used to play back virtual acoustic sources with an arbitrary loudspeaker setup [4]. To place a virtual source in the 2D plane, a pair of loudspeakers is required, while in the 3D case loudspeakers triplets are required. For each virtual source, a monophonic signal with different gains (dependent on the position of the virtual source) is fed to the selected loudspeakers from the full setup. The loudspeaker signals for all virtual sources are then summed up. VBAP applies a geometric approach to calculate the gains of the loudspeaker signals for the panning between the loudspeakers.

Another way to obtain 3D panning functions is presented in [6]. It is the extension of an existing 2D design as proposed by Poletti [7]. This approach uses a energy criterion to determine the panning functions.

Conventional Ambisonics decoding employs the commonly known mode matching process [8]. The modes are described by mode vectors that contain values of the spherical harmonics for a distinct direction of incidence. The combination of all directions given by the individual loudspeakers leads to the mode matrix of the loudspeaker setup. To reproduce the mode of a distinct source signal, the loudspeakers' modes are weighted in that way that the superpositioned modes of the individual loudspeakers sum up to the desired mode. To obtain the necessary weights, an inverse matrix representation of the loudspeaker mode matrix needs to be calculated. In terms of signal decoding the weights form the driving signal of the loudspeakers, and the inverse loudspeaker mode matrix is referred to as decoding matrix which is applied to the Ambisonics signal representation.

Another way to obtain the decoding matrix that is presented in this paper. It employs a process in a system estimation manner. Considering a set of possible directions of incidence, the panning functions related to the desired loudspeakers are calculated. This panning functions are now taken as output of an Ambisonics decoding process. The required input signal is the mode matrix of all considered directions. Therefore, the decoding matrix is obtained by right multiplying the weighting matrix by an inverse version of the mode matrix of input signals [7].

In the following, Section 2 reviews the VBAP approach for panning function calculation. Section 3 outlines the mode matching approach for Ambisonics decoding. In Section 4 the combination of VBAP and Ambisonics using the system estimation procedure is discussed. The informal listening test presented in Section 5 compares the different playback strategies. The major findings are summarised in the conclusions Section 6.

2. VECTOR BASE AMPLITUDE PANNING

Vector Base Amplitude Panning (VBAP) is used to place virtual acoustic sources with an arbitrary loudspeaker setup where the same distance of the loudspeakers from the listening position is assumed. VBAP uses three loudspeakers to place a virtual source in the 3D space. For each virtual source, a monophonic signal with different gains is fed to the loudspeakers to be used. The gains for the different loudspeakers are dependent on the position of the virtual source. VBAP is a geometric approach to calculate the gains of the loudspeaker signals for the panning between the loudspeakers. VBAP was proposed by Pulkki and is described in [4, 9]. We give a summary of the VBAP processing from the cited references.

In the 3D case, three loudspeakers arranged in a triangle build a vector base. Each vector base is identified by the loudspeaker numbers k,m,n and the loudspeaker position vectors $\boldsymbol{l}_k$, $\boldsymbol{l}_m$, $\boldsymbol{l}_n$ given in Cartesian coordinates normalised to unity length. The vector base for loudspeakers k,m,n is defined by

$$\boldsymbol{L}_{kmn} = \left(\begin{array}{c} \boldsymbol{l}_k, \boldsymbol{l}_m, \boldsymbol{l}_n \end{array} \right). \tag{1}$$

The desired direction $\Omega = (\theta, \phi)$ of the virtual source has to be given as azimuth angle ϕ and inclination angle θ. The unity length position vector $\boldsymbol{p}(\Omega)$ of the virtual source in Cartesian coordinates is therefore defined by

$$\boldsymbol{p}(\Omega) = \left(\begin{array}{c} \cos\phi \sin\theta, \sin\phi \sin\theta, \cos\theta \end{array} \right)^T. \tag{2}$$

A virtual source position can be represented with the vector base and the gain factors $\boldsymbol{g}(\Omega) = (\tilde{g}_k, \tilde{g}_m, \tilde{g}_n)^T$ by

$$\boldsymbol{p}(\Omega) = \boldsymbol{L}_{kmn}\boldsymbol{g}(\Omega) = \tilde{g}_k\boldsymbol{l}_k + \tilde{g}_m\boldsymbol{l}_m + \tilde{g}_n\boldsymbol{l}_n. \tag{3}$$

By inverting the vector base matrix the required gain factors can be computed by

$$\boldsymbol{g}(\Omega) = \boldsymbol{L}_{kmn}^{-1}\, \boldsymbol{p}(\Omega). \tag{4}$$

The vector base to be used is determined according to [4]: First the gains are calculated according to (4) for all vector bases. Then for each vector base the minimum over the gain factors is evaluated by $\tilde{g}_{\min} = \min\{\tilde{g}_k, \tilde{g}_m, \tilde{g}_n\}$. Finally the vector base where $\tilde{g}_{\min}$ has the highest value is used. The resulting gain factors must not be negative. Depending on the listening room acoustics the gain factors may be normalised for energy preservation.

3. AMBISONICS

The Ambisonics representation is a sound field description method employing a mathematical approximation of the sound field in one location. Using the spherical coordinate system, the pressure at point $\boldsymbol{r} = (r, \theta, \phi)$ in space is described by means of the spherical Fourier transform [8, 10]

$$p(\boldsymbol{r}, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} A_n^m(k)\, j_n(kr)\, Y_n^m(\theta, \phi), \tag{5}$$

where k is the wave number. Normally n runs to a finite order M. The coefficients $A_n^m(k)$ of the series describe the sound field (assuming sources outside the region of validity [10]), $j_n(kr)$ is the spherical Bessel function of first kind and $Y_n^m(\theta, \phi)$ denote the spherical harmonics. Coefficients $A_n^m(k)$ are regarded as Ambisonics coefficients in this context. The spherical harmonics $Y_n^m(\theta, \phi)$ only depend on the inclination and azimuth angles and describe a function on the unity sphere [10].

For reasons of simplicity often plain waves are assumed for sound field reproduction. The Ambisonics coefficients describing a plane wave as an acoustic source from direction Ω_s are [8]

$$A_{n,\text{plane}}^m(\Omega_s) = 4\pi\, \mathrm{i}^n\, Y_n^m(\Omega_s)^*. \tag{6}$$

Their dependency on wave number k decreases to a pure directional dependency in this special case. For a limited order M the coefficients form a vector $\boldsymbol{A}$ that may be arranged as

$$\boldsymbol{A}(\Omega_s) = \left[\begin{array}{cccccc} A_0^0 & A_1^{-1} & A_1^0 & A_1^1 & \ldots & A_M^M \end{array} \right]^T \tag{7}$$

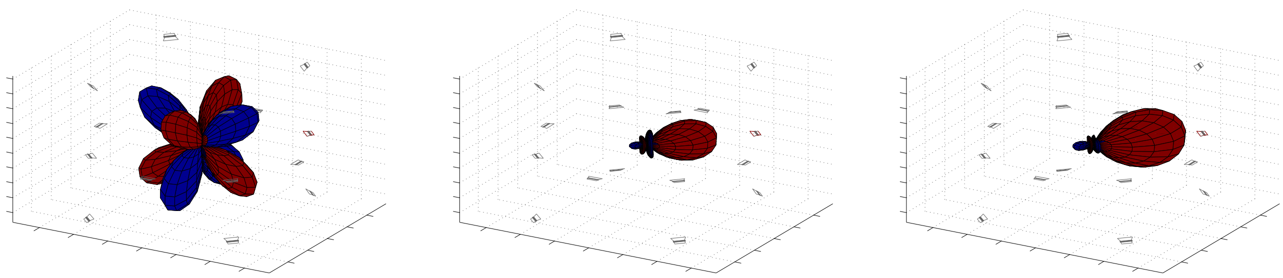
holding $O = (M+1)^2$ elements. The same arrangement is used for the spherical harmonics coefficients yielding a vector $\boldsymbol{Y}(\Omega_s)^* = \left[\begin{array}{cccccc} Y_0^0 & Y_1^{-1} & Y_1^0 & Y_1^1 & \ldots & A_M^M \end{array} \right]^H$. Superscript H denotes the complex conjugate transpose.

To calculate loudspeaker signals from an Ambisonics representation of a sound field, mode matching is a commonly used approach [8]. The basic idea is to express a given Ambisonics sound field description $\boldsymbol{A}(\Omega_s)$ by a weighted sum of the loudspeakers' sound field descriptions $\boldsymbol{A}(\Omega_l)$

$$\boldsymbol{A}(\Omega_s) = \sum_{l=1}^{L} w_l \boldsymbol{A}(\Omega_l) \tag{8}$$

where Ω_l denote the loudspeakers' directions, w_l are weights, and L is the number of loudspeakers. To derive panning functions from (8), we assume a known direction of incidence Ω_s. If source and speaker sound fields are both plane waves, the factor $4\pi\mathrm{i}^n$ (see (6)) can be dropped and (8) only depends on the complex conjugates of spherical harmonic vectors, also referred to as modes. Using matrix notation this is written as

$$\boldsymbol{Y}(\Omega_s)^* = \boldsymbol{\Psi}\, \boldsymbol{w}(\Omega_s), \tag{9}$$



(a) This beampattern is derived from the non regularised mode matching case.

(b) The same as in Figure 2(a), but with a regularised mode matrix.

(c) The beampattern of the VBAP-derived Ambisonics decoding matrix shows smaller side lobes.

Figure 2: Panning functions obtained for loudspeaker #2 in Figure 1. Red means a positive amplitude, blue a negative amplitude. All examples use a decoding matrix of order $M = 3$.

where $\boldsymbol{\Psi}$ is the mode matrix of the loudspeaker setup

$$\boldsymbol{\Psi} = [\boldsymbol{Y}(\Omega_1)^*, \boldsymbol{Y}(\Omega_2)^*, \ldots, \boldsymbol{Y}(\Omega_L)^*] \tag{10}$$

with $O \times L$ elements. To obtain the desired weighting vector $\boldsymbol{w}$, several strategies to accomplish this are discussed in [8] or [5]. If $M = 3$ is chosen, $\boldsymbol{\Psi}$ is square and may be invertible. Due to the irregular loudspeaker setup the matrix is badly scaled, though. In such a case, often the pseudo inverse matrix is chosen and

$$\boldsymbol{D} = \left[\boldsymbol{\Psi}^H \boldsymbol{\Psi}\right]^{-1} \boldsymbol{\Psi}^H, \tag{11}$$

yields a $L \times O$ decoding matrix $\boldsymbol{D}$. Finally we can write

$$\boldsymbol{w}(\Omega_s) = \boldsymbol{D}\boldsymbol{Y}(\Omega_s)^*, \tag{12}$$

where the weights $\boldsymbol{w}(\Omega_s)$ are the minimum energy solution for (9). The consequences from using the pseudo inverse are described in the next section.

4. AMBISONICS USING VBAP

This section describes the link between panning functions and the Ambisonics decoding matrix. Starting with Ambisonics, the panning functions for the individual loudspeakers can be calculated using (12). Let

$$\boldsymbol{\Xi} = [\boldsymbol{Y}(\Omega_1)^*, \boldsymbol{Y}(\Omega_2)^*, \ldots, \boldsymbol{Y}(\Omega_S)^*] \tag{13}$$

qbe the mode matrix of S input signal directions (Ω_s) (e. g. a spherical grid with an inclination angle running in steps of one degree from $1 \ldots 180$ and an azimuth angle from $1 \ldots 360$ respectively). This mode matrix has $O \times S$ elements. Using (12), the resulting matrix $\boldsymbol{W}$ has $L \times S$ elements, row l holds the S panning weights for the respective loudspeaker:

$$\boldsymbol{W} = \boldsymbol{D}\boldsymbol{\Xi}. \tag{14}$$

As a representative example the panning function of loudspeaker #2 is shown as beam pattern in Figure 2(a). The decode matrix $\boldsymbol{D}$ is order $M = 3$ in this example. It is easily recognised that the panning function values do not refer to the physical positioning of the loudspeaker at all. This is due to the mathematical irregular positioning of the loudspeakers which is not sufficient as a spatial sampling scheme for the chosen order. This problem can be overcome by regularisation of the loudspeaker mode matrix $\boldsymbol{\Psi}$ in (11). This solution works at the expense of spatial resolution of the decoding matrix which in turn may be expressed as a lower Ambisonics order. Figure 2(b) shows an example using the mean of eigenvalues of the mode matrix for regularisation. The direction of the addressed loudspeaker is now clearly recognised.

As outlined in the introduction, another way to obtain a decoding matrix $\boldsymbol{D}$ for playback of Ambisonics signals is possible when the panning functions are already known [7]. The panning functions $\boldsymbol{W}$ are viewed as desired signal defined on a set of virtual source directions Ω, and the mode matrix $\boldsymbol{\Xi}$ (13) of these directions serves as input signal. Then the decoding matrix can be calculated using

$$\boldsymbol{D} = \boldsymbol{W}\boldsymbol{\Xi}^H \left[\boldsymbol{\Xi}\boldsymbol{\Xi}^H\right]^{-1}, \tag{15}$$

where the latter term is the pseudo inverse of the mode matrix $\boldsymbol{\Xi}$ [7]. In our new approach we take the panning functions in $\boldsymbol{W}$ from VBAP and calculate an Ambisonics decoding matrix from this.

The panning functions for $\boldsymbol{W}$ are taken as gain values $\boldsymbol{g}(\Omega)$ calculated using (4), where Ω is chosen according to (13). The resulting decode matrix using (15) is an Ambisonics decoding matrix facilitating the VBAP panning functions. An example is shown in Figure 2(c). The side lobes are significantly smaller than the side lobes of the regularised mode matching result. Moreover, the VBAP-derived beam pattern for the individual loudspeakers follow the geometry of the loudspeaker setup as the VBAP panning functions depend on the vector base of the addressed direction. As a consequence, this new approach produces better results over all directions of the loudspeaker setup.

5. LISTENING TEST

For the evaluation of the localisation of a single source, we compare a virtual source against a real source as a reference in a listening test. The listening test is similar to the one in [11], where the Ambisonics decoding with 3D robust panning is evaluated. For the real source we use a loudspeaker at the desired position. The playback methods used are VBAP, Ambisonics mode matching decoding, Ambisonics decoding using the 3D robust panning approach [6], and the newly proposed Ambisonics decoding using VBAP panning functions. For the latter three methods, for each tested position and each tested input signal, an Ambisonics signal of third order is generated. This synthetic Ambisonics signal is then decoded using the corresponding decoding matrices. The test signals used are broadband pink noise and a male speech signal. The tested positions are placed in the frontal region with the directions

$$\Omega_1 = (76.1^\circ, -23.2^\circ), \qquad \Omega_2 = (63.3^\circ, -4.3^\circ). \tag{16}$$

The listening test was conducted in an acoustic room with a mean reverberation time of approximately 0.2 s. Nine people participated in the listening test. The test subjects were asked to grade the spatial playback performance of all playback methods compared to the reference. A single grade value had to be found to represent the localisation of the virtual source and timbre alterations. Figure 3 shows the listening test results.

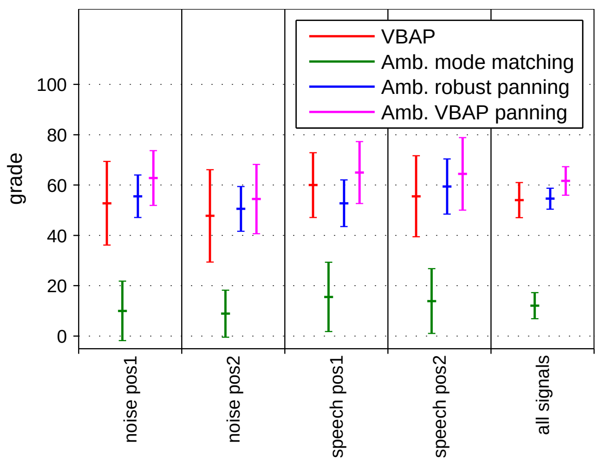


Figure 3: Listening test results as grades on a scale from 0 to 100 where 100 corresponds to excellent quality. Mean values and 95% confidence intervals are shown for the considered playback methods.

The unregularised Ambisonics mode matching decoding is graded worse than the other methods under test. This result corresponds to Figure 2. Due to the bad localisation properties of the Ambisonics mode matching, this method serves as anchor in this listening test. Please note that the mode matching approach can be improved with a regularisation for the calculation of the pseudo inverse as shown in Figure 2(b). In general, the three other methods show similar grades, where for the noise signal the confidence intervals are greater for VBAP. The mean values show the highest values for the Ambisonics decoding using VBAP panning functions. Thus, although the spatial resolution is reduced – due to the Ambisonics order used – this method shows advantages over the parametric VBAP approach. Compared to VBAP, both Ambisonics decoding with robust and VBAP panning functions have the advantage that not only three loudspeakers are used to render the virtual source. In VBAP single loudspeakers may be dominant if the virtual source position is close to one of the physical positions of the loudspeakers.

Most subjects reported less timbre alterations for the Ambisonics driven VBAP than for directly applied VBAP. The problem of timbre alterations for VBAP is already known from Pulkki [9]. In opposite to VBAP, our method uses more than three loudspeakers for playback as stated above, but surprisingly produces less colouration of the signal. This finding is a topic of further research.

6. CONCLUSIONS

In this contribution we presented a way to obtain the Ambisonics decoding matrix from the VBAP panning functions. For different loudspeaker setups, this approach is compared to matrices of the mode matching approach. Properties and consequences of these decoding matrices are discussed.

In summary, the newly proposed Ambisonics decoding with VBAP panning functions avoids typical problems of the well known mode matching approach. An informal listening test has shown that VBAP driven Ambisonics decoding can produce a spatial playback quality better than the direct used VBAP can produce. The proposed method requires only a sound field description while VBAP requires a parametric description of the virtual sources to be rendered.

References

[1] K. Hamasaki, T. Nishiguchi, R. Okumaura, and Y. Nakayama, “Wide listening area with exceptional spatial sound qualtity of a 22.2 multichannel sound system,” in *Audio Engineering Society Preprints*, Vienna, Austria, May 2007, paper 7037 presented at the 122nd Convention.

[2] “Mdg-musikproduktion dabringhaus und grimm,” www.mdg.de.

[3] T. Holman, *Sound for Film and Television*, 2nd ed. Boston: Focal Press, 2002.

[4] V. Pulkki, “Virtual sound source positioning using vector base amplitude panning,” *J. Audio Eng. Soc.*, vol. 45, no. 6, pp. 456–466, June 1997.

[5] H. Pomberger and F. Zotter, “An ambisonics format for flexible playback layouts,” in *Proceedings of the 1st Ambisonics Symposium*, Graz, Austria, July 2009.

[6] J.-M. Batke and F. Keiler, “Investigation of robust panning functions for 3D loudspeaker setups,” in *128th Convention of the Audio Eng. Soc.*, London, UK, May 2010.

[7] M. Poletti, “Robust two-dimensional surround sound reproduction for nonuniform loudspeaker layouts,” *J. Audio Eng. Soc.*, vol. 55, no. 7/8, pp. 598–610, July/August 2007.

[8] M. A. Poletti, “Three-dimensional surround sound systems based on spherical harmonics,” *J. Audio Eng. Soc.*, vol. 53, no. 11, pp. 1004–1025, Nov. 2005.

[9] V. Pulkki, “Spatial sound generation and perception by amplitude panning techniques,” Ph.D. dissertation, Helsinki University of Technology, 2001. [Online]. Available: http://lib.tkk.fi/Diss/2001/isbn9512255324/

[10] E. G. Williams, *Fourier Acoustics*. Academic Press, 1999.

[11] F. Keiler and J.-M. Batke, “Evaluation of virtual source localisation using 3D loudspeaker setups,” in *128th Convention of the Audio Eng. Soc.*, London, UK, May 2010.